Autonomous Vocal and Backing Track Mixing

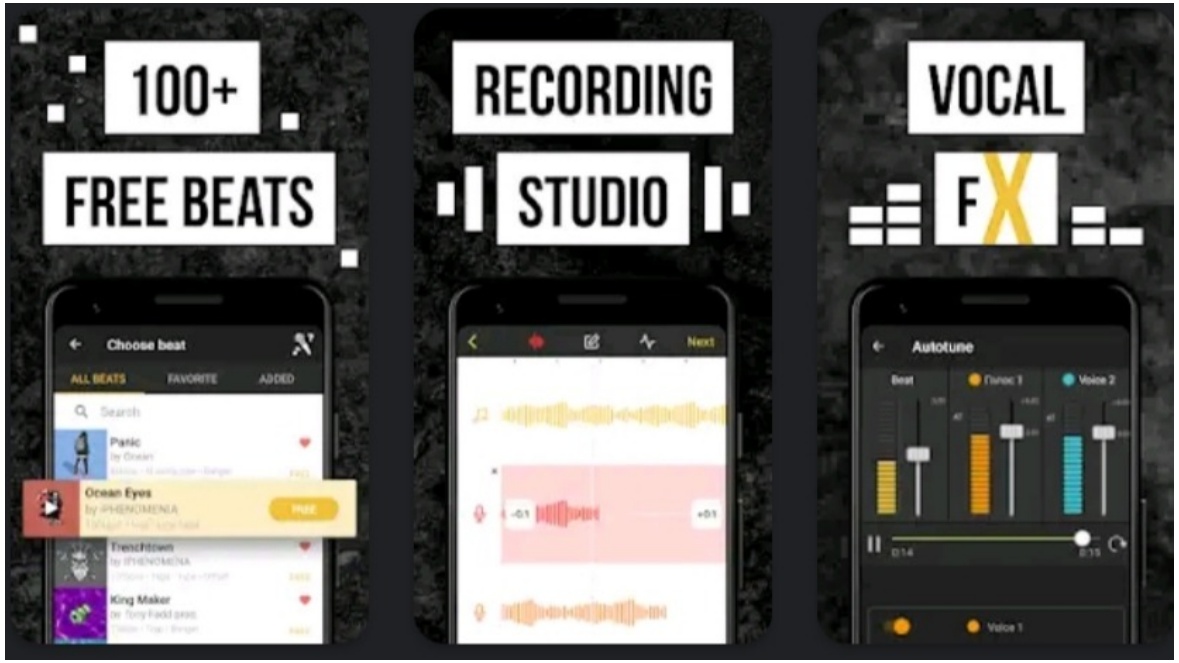
Kelian (Mike) Li

Music Informatics Group

Georgia Tech Center for Music Technology
College of Design

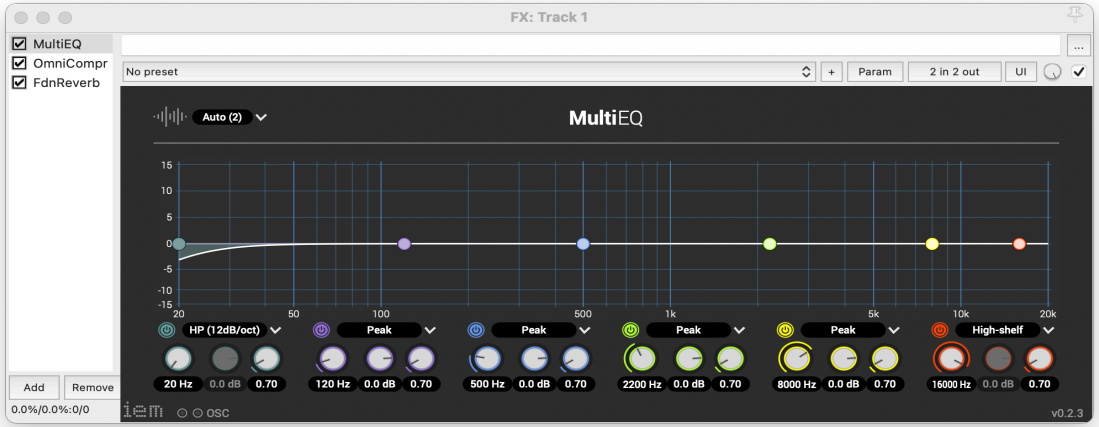
Motivation

- Karaoke apps
- Amateur music makers



Audio Processing

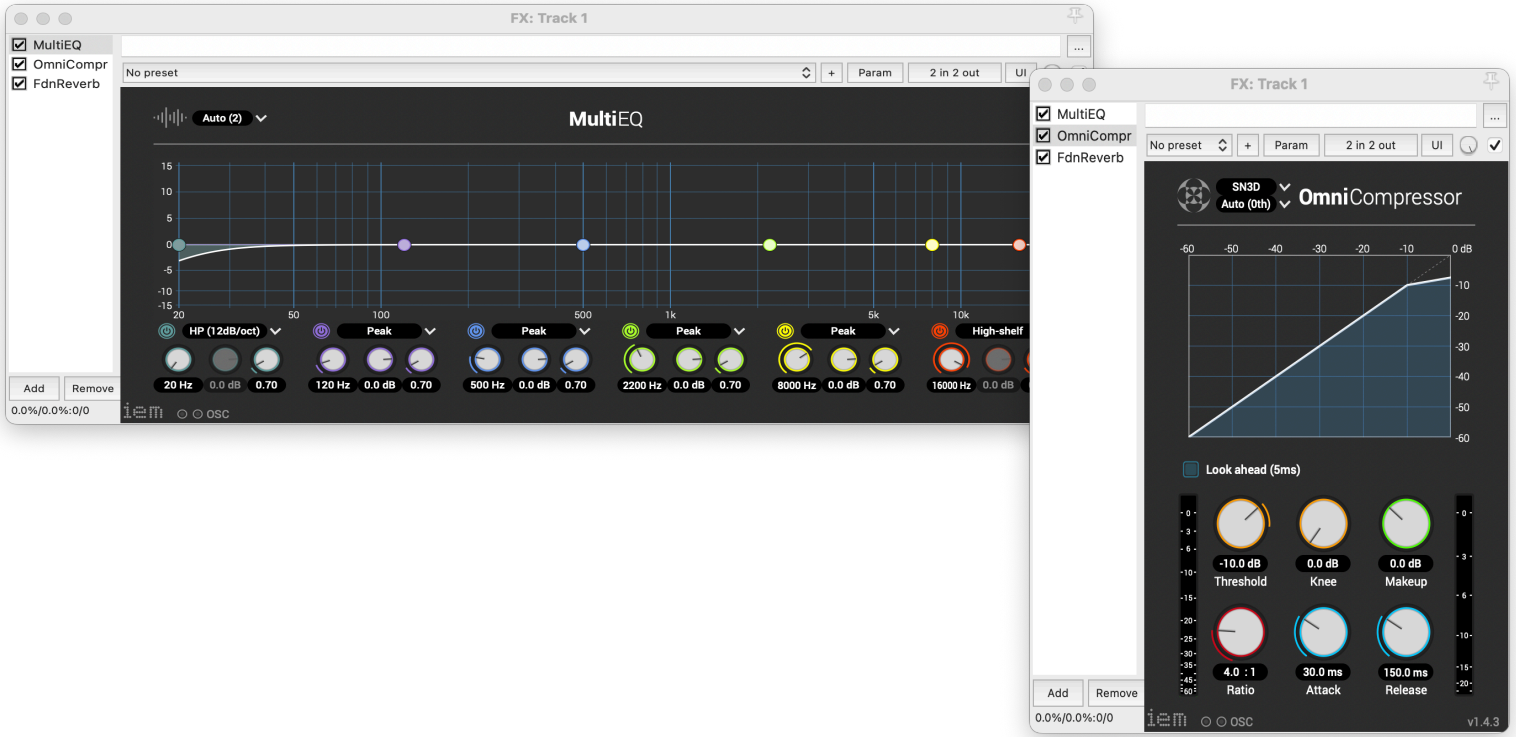
- EQ



Georgia Tech Center for Music Technology
College of Design

Audio Processing

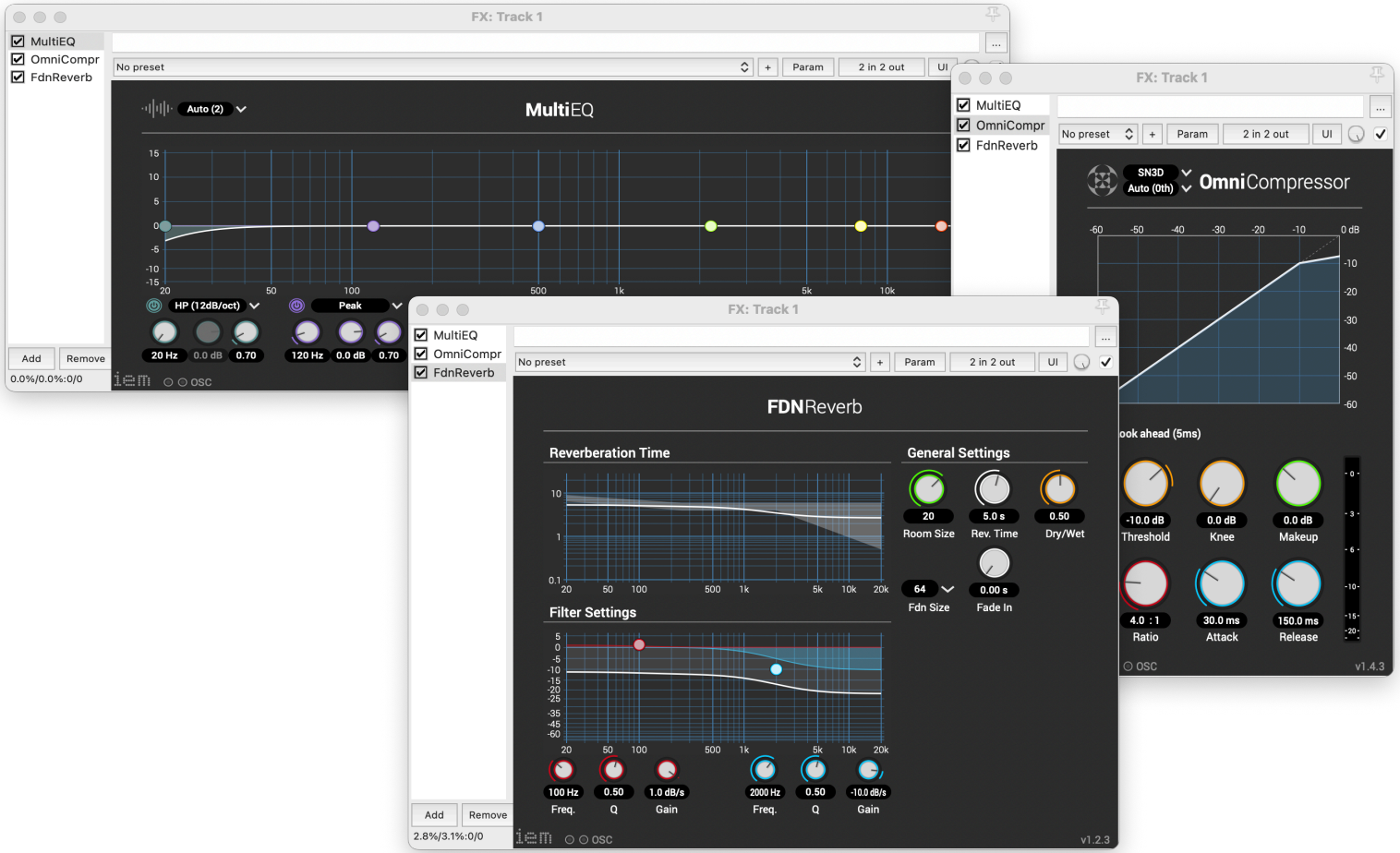
- EQ
- Compression



Georgia Tech Center for Music Technology
College of Design

Audio Processing

- EQ
- Compression
- Reverb



Georgia Tech Center for Music Technology
College of Design

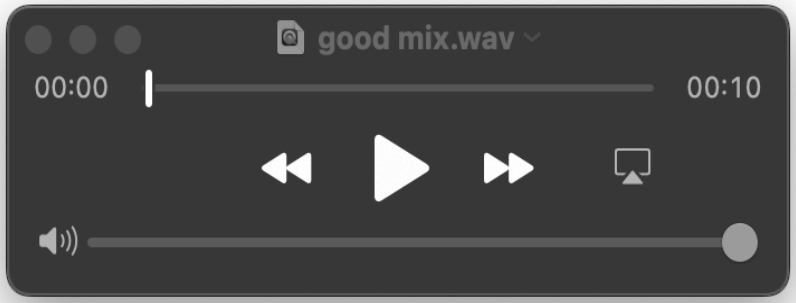
Audio Processing

- EQ
- Compression
- Reverb
- Level Balance

Georgia Tech Center for Music Technology
College of Design

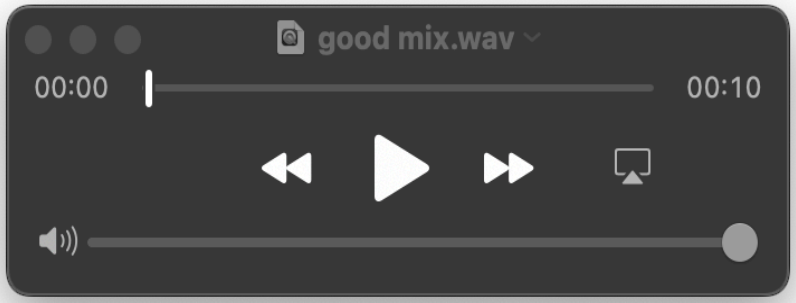
Audio Example

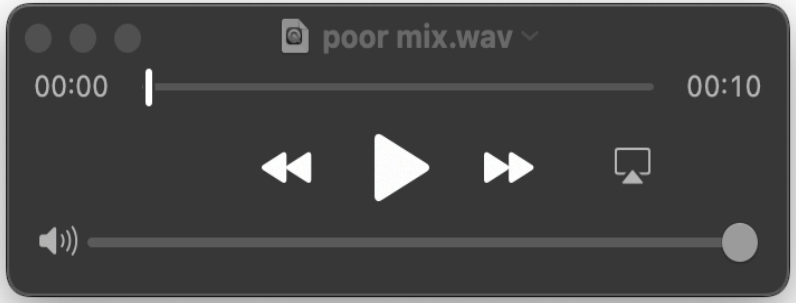
- “Very Good” mix



Audio Example

- “Very Good” mix
- “Very Poor” mix





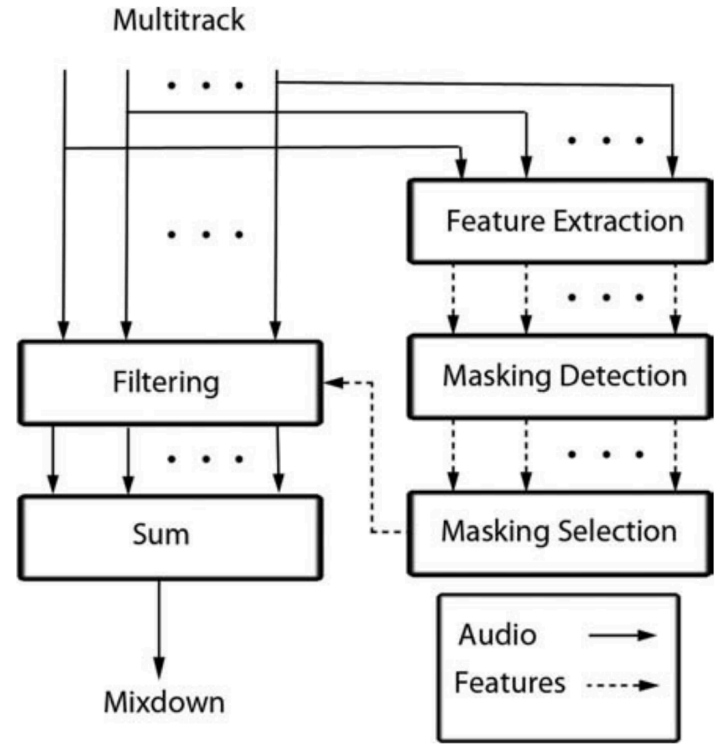
Data Analysis Approach: level and compression

Use the average values extracted from the source-separated Million Song Dataset

- Level balance
  - -1.77 dB vocal-to-backing track ratio
- Compression
  - 16.4 dB loudness range

Data Analysis Approach: EQ

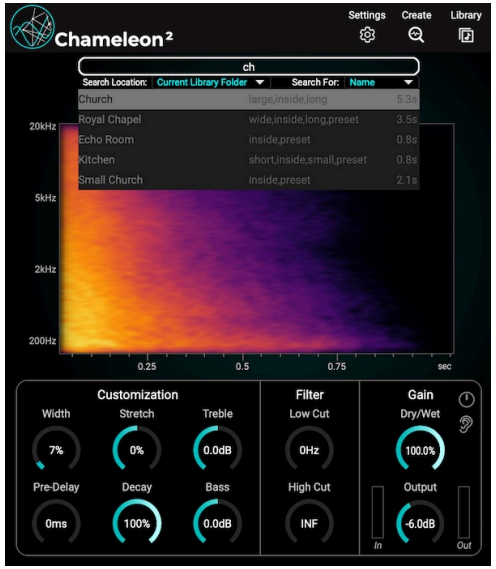
- EQ
  - Frequency unmasking



Data Analysis Approach: reverb

1. Get the estimated impulse responses from the Chameleon plugin
2. Estimate the reverb parameters by the genetic algorithm
3. Use mean values extracted from the MUSDB18 train set

- **Reverb**
  - Dry/wet ratio: 11.5%
  - Reverb time: Linear mapping from tempo
  - Room size: 14.54
  - Fade in time: 0.68 s



Georgia Tech Center for Music Technology
College of Design

Deep Learning Approach

Train a convolutional neural network to predict direct or intermediate mixing parameters based on the input audio

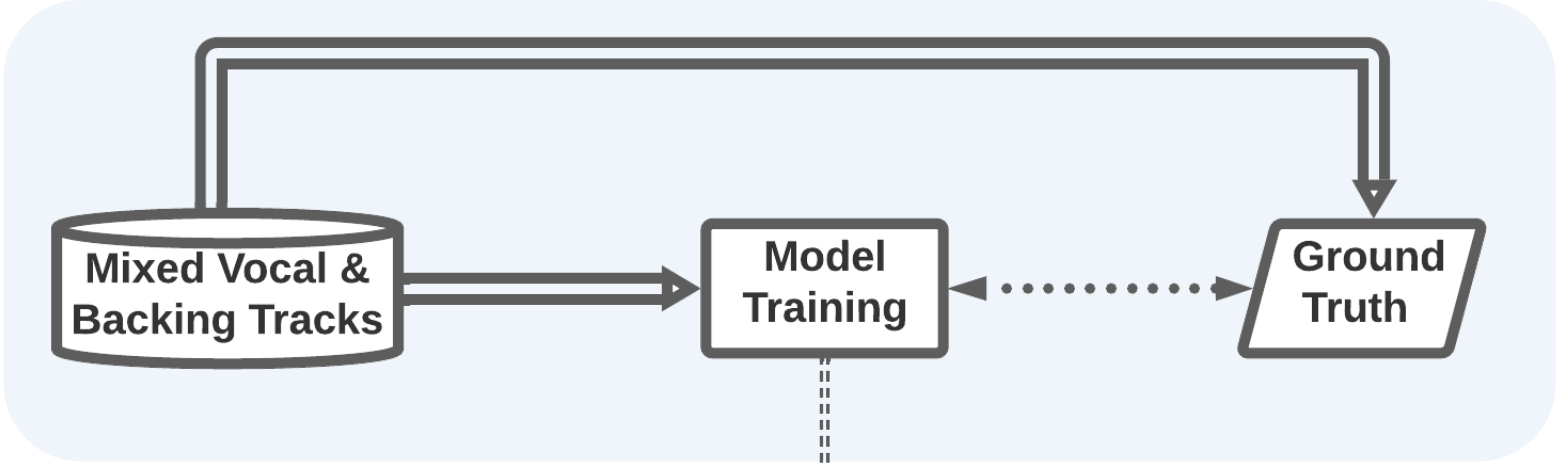
- model input:

  Mel-spectrogram of the vocal and the backing track

- model output:

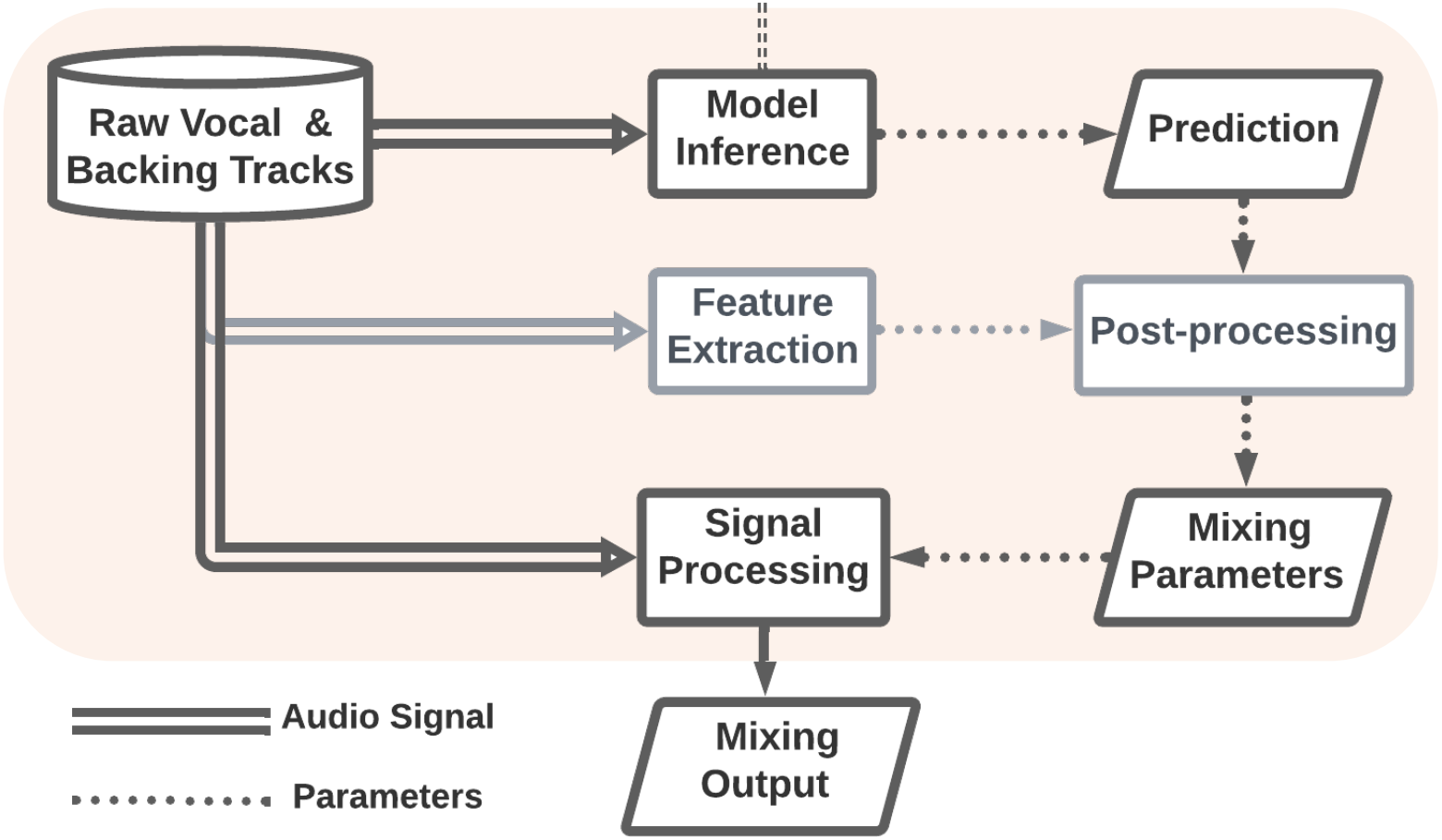
  the same as the previous approach

Model Training



Ground truth is the direct or intermediate mixing parameters

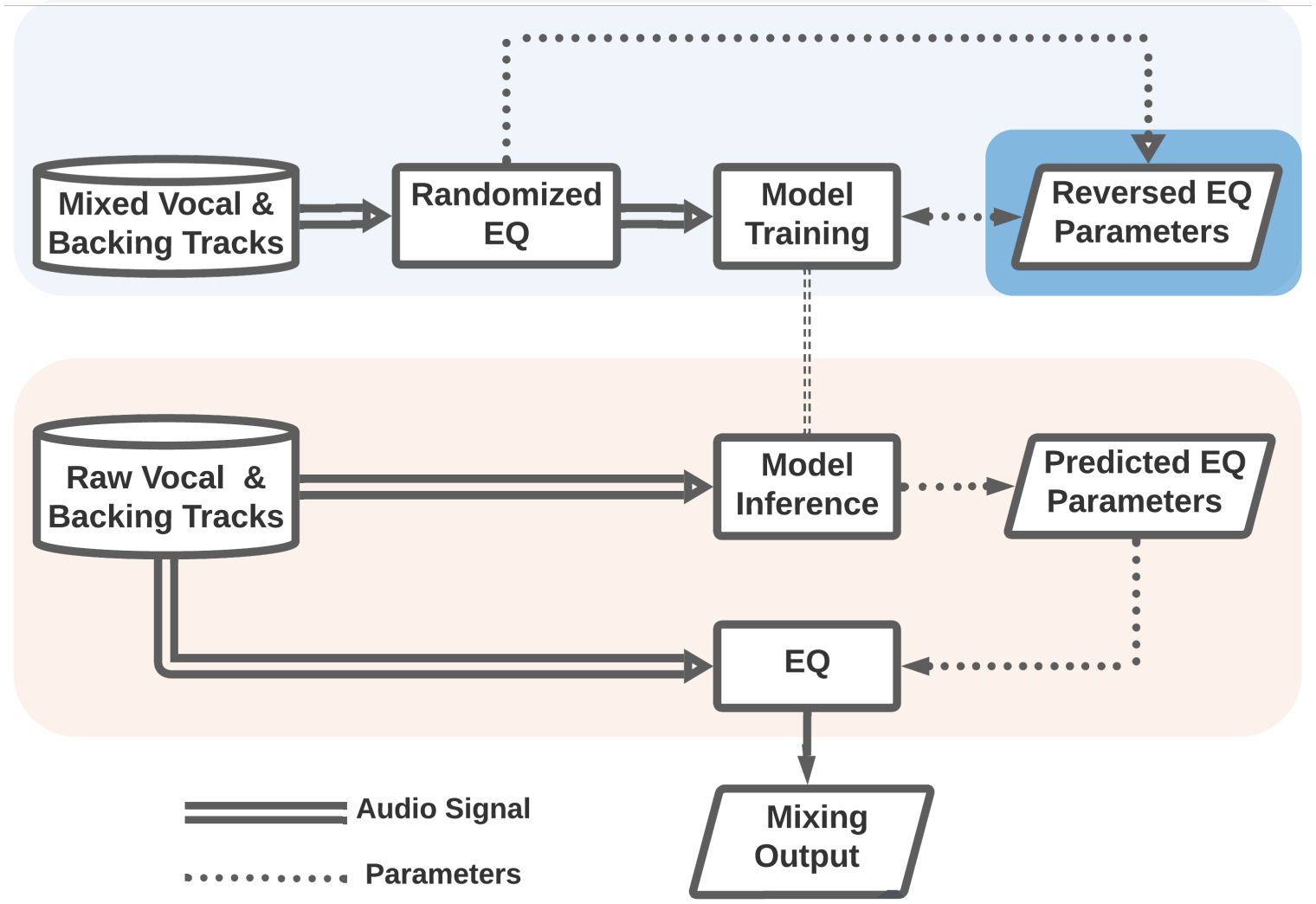
Model Inference



Deep Learning Approach: EQ

Model Training

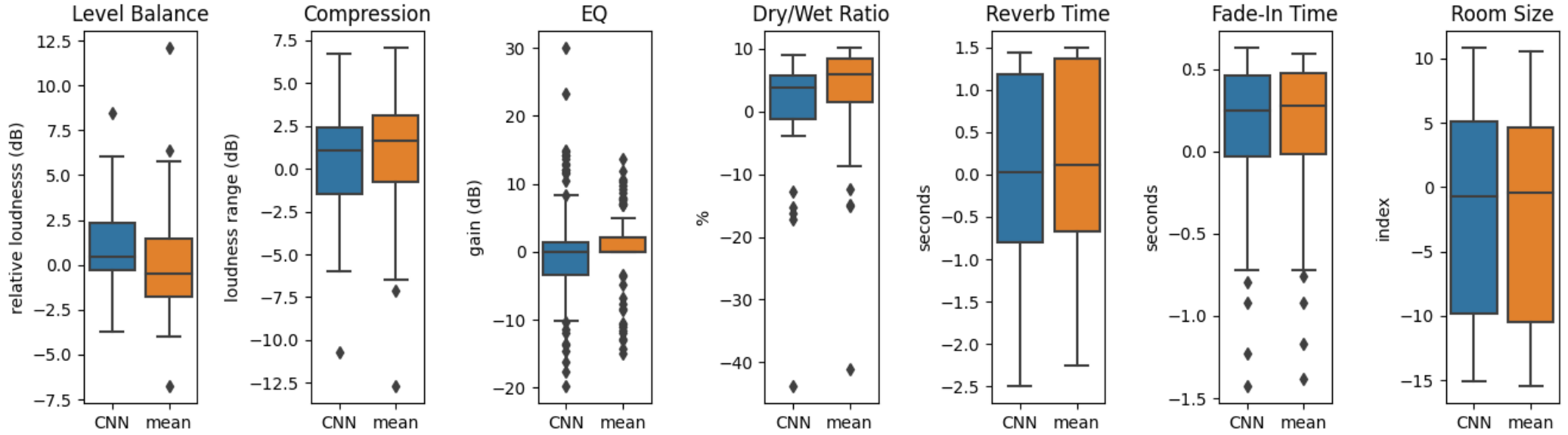
Model Inference



If the mixed vocal is boosted at a center frequency, we should learn to cut at that frequency.

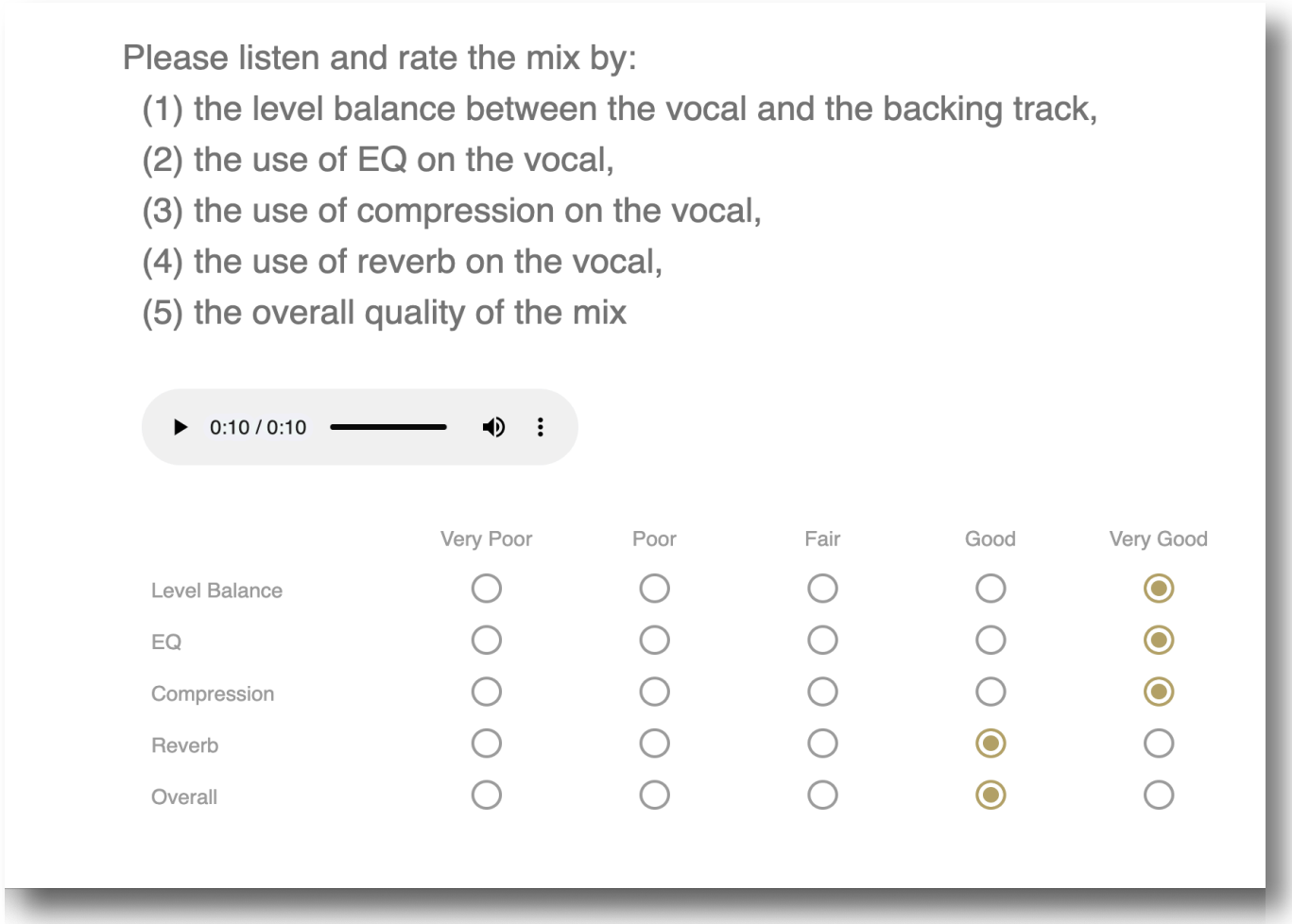
Objective Evaluation

Validation error on the MUSDB test set, 48 songs in total



	relative loudness	loudness range	EQ gain	dry/wet ratio	reverb time	FT	RS	LCF	LCG	LCQ	HCF	HCG	HCQ
CNN	**1.64**	**2.63**	4.48	**6.13**	**1.006**	**0.401**	**7.30**	41.55	**3.62**	**0.355**	**3390**	**5.01**	**0.14**
mean	2.13	2.88	**3.33**	7.16	1.007	0.403	7.31	**40.26**	3.65	0.360	3462	5.35	0.17

The Final Step: listening test



https://t.ly/0kur

Georgia Tech Center for Music Technology
College of Design